

Article

Complex Correntropy with Variable Center: Definition, Properties, and Application to Adaptive Filtering

Fei Dong, Guobing Qian *  and Shiyuan Wang 

College of Electronic and Information Engineering, Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing, Southwest University, Chongqing 400715, China; doufeiii@126.com (F.D.); wsy@swu.edu.cn (S.W.)

* Correspondence: guobingqian@swu.edu.cn

Received: 14 November 2019; Accepted: 4 January 2020; Published: 6 January 2020



Abstract: The complex correntropy has been successfully applied to complex domain adaptive filtering, and the corresponding maximum complex correntropy criterion (MCCC) algorithm has been proved to be robust to non-Gaussian noises. However, the kernel function of the complex correntropy is usually limited to a Gaussian function whose center is zero. In order to improve the performance of MCCC in a non-zero mean noise environment, we firstly define a complex correntropy with variable center and provide its probability explanation. Then, we propose a maximum complex correntropy criterion with variable center (MCCC-VC), and apply it to the complex domain adaptive filtering. Next, we use the gradient descent approach to search the minimum of the cost function. We also propose a feasible method to optimize the center and the kernel width of MCCC-VC. It is very important that we further provide the bound for the learning rate and derive the theoretical value of the steady-state excess mean square error (EMSE). Finally, we perform some simulations to show the validity of the theoretical steady-state EMSE and the better performance of MCCC-VC.

Keywords: complex; MCCC-VC; variable center; stability; EMSE

1. Introduction

Choosing the appropriate cost function (usually the statistical measure of error signal) is the key problem in adaptive filtering theory and application [1–3]. In the presence of Gaussian noise, it is best to using the minimum mean square error (MMSE) criterion. Therefore, a series of MMSE based algorithms [4–7] have emerged during the past decades. The MMSE based algorithms use the mean square value of the error between the desired signal and output signal as the cost function, which has many attractive features, such as convexity and smoothness. In addition, MMSE has low computational complexity since it only needs to calculate the second order statistics of the signals. However, in many non-Gaussian cases, the MMSE based algorithms are not robust. To improve this shortcoming, many kinds of non-MMSE criteria based algorithms have been developed in [8–16]. Since signals are often expressed in complex forms in many practical scenarios [17,18], adaptive filtering in complex domain is of great significance. During the past few years, some information criteria based algorithms have been proposed for complex domain adaptive filtering [19–22]. Particularly recently Guimarães defined a new similarity measurement between two complex variables based on complex correntropy [19,20] and proposed the maximum complex correntropy criterion (MCCC) algorithm. MCCC uses a complex Gaussian function as the kernel function, and derives the updatation of weight based on Wirtinger Calculus. The complex Gaussian kernel function is desirable due to its smoothness and strict positive-definiteness. The performance of the MCCC algorithm is better than classic MMSE

based algorithms, and is robust to non-Gaussian noise. Moreover, MCCC has been widely applied to many fields of machine learning and signal processing [23,24].

According to the MCCC, given two complex variables $C_1 = A_1 + jB_1$ and $C_2 = A_2 + jB_2$, complex correntropy is defined by [19,20]

$$V_{\sigma}^C(C_1, C_2) = E[\kappa(C_1 - C_2)] \quad (1)$$

where A_1, B_1, A_2, B_2 represent real variables, $E[\cdot]$ denotes the expectation, and $\kappa(C_1 - C_2)$ denotes the kernel function with

$$\begin{aligned} \kappa(C_1 - C_2) &= G_{\sigma}^C(C_1 - C_2) \\ &= \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(C_1 - C_2)(C_1 - C_2)^*}{2\sigma^2}\right) \end{aligned} \quad (2)$$

and $\sigma > 0$ is the kernel width.

The purpose of adaptive filtering is to estimate the target variable T in some sense by designing a model M to construct a output Y from input X . Under MCCC, we find this mode by maximizing the complex correntropy between T and Y :

$$M^* = \underset{M \in \mathcal{M}}{\operatorname{argmax}} V_{\sigma}^C(T, Y) = \underset{M \in \mathcal{M}}{\operatorname{argmax}} E[G_{\sigma}^C(T - Y)] \quad (3)$$

where \mathcal{M} is the model assumption space which contains the possible models to construct the output Y from input X , and M^* is the optimal model.

However, the center of complex correntropy is always at zero, which is not the best option in the case of non-zero mean noise. Although the maximum correntropy criterion with variable center in [25] and [26] can be suitable for the variable center, they cannot be used for complex domain adaptive filtering. To overcome their defects, this paper proposes the maximum complex correntropy criterion with variable center (MCCC-VC).

The main contributions of this research lie in the following aspects: (1) we define a MCCC-VC and give its probability explanation; (2) based on the MCCC-VC, we propose a novel adaptive filtering algorithm in complex domain by utilizing the gradient descend approach; (3) we give effective and feasible methods to estimate the kernel center and update the kernel width adaptively; (4) we derive the bound for the learning rate, and the theoretical steady-state excess mean square error (EMSE) of the MCCC-VC algorithm, and verify the theoretical analysis by simulations.

The organization of this paper is as follows: Section 2 defines the complex correntropy with variable center and studies its properties. Section 3 proposes the MCCC-VC algorithm and provides the method for the optimization of the parameters. In addition, Section 3 also studies the convergence of the MCCC-VC algorithm and derives the theoretical steady-state excess mean square error (EMSE). Section 4 verifies the correctness of the theoretical conclusions and the superior performance of the MCCC-VC algorithm. Finally, Section 5 summarizes the conclusion of this paper.

2. Complex Correntropy with Variable Center

For two complex variables, the target variable T and the output Y , the complex correntropy with variable center is defined as:

$$V_{\sigma,c}^C(T, Y) = E[G_{\sigma}^C(T - Y - c)] = E\left[\frac{1}{2\pi\sigma^2} \exp\left(-\frac{(T - Y - c)(T - Y - c)^*}{2\sigma^2}\right)\right] \quad (4)$$

where $c \in \mathbb{C}$ represents the center of the kernel function. When $c = 0$, (4) will return to the original complex correntropy.

The complex correntropy with variable center c consists of the whole even moments of $T - Y$ about the center c , which is as follows:

$$V_{\sigma,c}^C(T, Y) = \frac{1}{2\pi\sigma^2} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n n!} E \left[\frac{|e - c|^{2n}}{\sigma^{2n}} \right] \quad (5)$$

where $e = T - Y$ is the complex valued error variable. With the increase of σ , the higher-order moments around the variable center c would attenuate quickly. Therefore, the second-order moment is the key factor which determines the value. In particular, when $c = E[e]$ and $\sigma \rightarrow \infty$, maximizing the complex correntropy with c would be equal to minimizing the variance of the error.

Moreover, when $\sigma \rightarrow 0$, we obtain

$$\begin{aligned} \lim_{\sigma \rightarrow 0} V_{\sigma,c}^C(T - Y) &= \lim_{\sigma \rightarrow 0} \iint \iint G_{\sigma}^C(t_R - y_R - c_R, t_I - y_I - c_I) p_{TY}(t_R, t_I, y_R, y_I) dt_R dt_I dy_R dy_I \\ &= \iint \iint \delta(t_R - y_R - c_R, t_I - y_I - c_I) p_{TY}(t_R, t_I, y_R, y_I) dt_R dt_I dy_R dy_I \\ &= \iint p_{TY}(t_R, t_I, t_R - c_R, t_I - c_I) dt_R dt_I \end{aligned} \quad (6)$$

where $\delta(x, y)$ is the two-dimensional Dirac function with $\begin{cases} \iint \delta(x, y) dx dy = 1 \\ \delta(x, y) = 0, x^2 + y^2 \neq 0 \end{cases}$, the second line is derived based on the fact that $\lim_{\sigma \rightarrow 0} G_{\sigma}^C(x, y) = \lim_{\sigma \rightarrow 0} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$ has the same property as $\delta(x, y)$, t_R , y_R , and c_R are the real parts of t , y and c , t_I , y_I , and c_I are the imaginary parts of t , y , and c , and $p_{TY}(t_R, t_I, y_R, y_I)$ denotes the joint probability density function (PDF) of (T, Y) . Furthermore, we derive the following result:

$$\begin{aligned} \lim_{\sigma \rightarrow 0} V_{\sigma,c}^C(T - Y) &= \lim_{\sigma \rightarrow 0} \iint G_{\sigma}^C(\varepsilon_R - c_R, \varepsilon_I - c_I) p_e(\varepsilon_R, \varepsilon_I) d\varepsilon_R d\varepsilon_I \\ &= \iint \delta(\varepsilon_R - c_R, \varepsilon_I - c_I) p_e(\varepsilon_R, \varepsilon_I) d\varepsilon_R d\varepsilon_I \\ &= \iint p_e(c_R, c_I) d\varepsilon_R d\varepsilon_I \end{aligned} \quad (7)$$

where $p_e(\varepsilon_R, \varepsilon_I)$ is the joint PDF of error. Thus, when $\sigma \rightarrow 0$, the value of complex correntropy with variable center c would approach $p_e(\varepsilon_R, \varepsilon_I)$ evaluated at (c_R, c_I) .

3. MCCC-VC Algorithm

In this part, we derive a novel adaptive filtering algorithm based on maximum complex correntropy criterion (i.e., minimum complex correntropy loss) with variable center (MCCC-VC).

3.1. Cost Function

We apply the MCCC-VC to adaptive filtering and derive the cost function as follows:

$$\begin{aligned} J_{VC-loss}^C &= G_{\sigma}^C(0) - E[G_{\sigma}^C(e(k) - c(k))] \\ &= \frac{1}{2\pi\sigma^2} \left\{ 1 - E \left[\exp \left[-\frac{((e(k) - c(k))(e(k) - c(k))^*)}{2\sigma^2} \right] \right] \right\} \end{aligned} \quad (8)$$

where

$$e(k) = d(k) - \mathbf{w}^H \mathbf{x}(k) \quad (9)$$

is the error at time instant k , $\mathbf{w} = [w_1 \ w_2 \ \cdots \ w_m]^T$ is the filter weight, $d(k)$ is the desired signal at time instant k , $\mathbf{x}(k) = [x(k) \ x(k-1) \ \cdots \ x(k-m+1)]^T$ is the input signal at time instant k , $c(k)$ is the center of the kernel at time instant k .

The essential idea behind the cost function (8) is that, in the practical case, even when the error distribution is non-zero-mean, the proposed MCCC-VC can perform well, because MCCC-VC matches well the error distribution.

Figure 1 compares the surfaces of the proposed MCCC-VC with MCCC, where the noise is non-zero-mean complex Gaussian noise with unit variance. For visualization, we chose $m = 1$, and set the system parameter and the mean of the noise as $\mathbf{w}_0 = 5 + 5i$ and $c = 6 + 6i$, respectively. One can see that the cost function of MCCC-VC is minimized at \mathbf{w}_0 , whereas the cost function of MCCC is minimized at some other place.

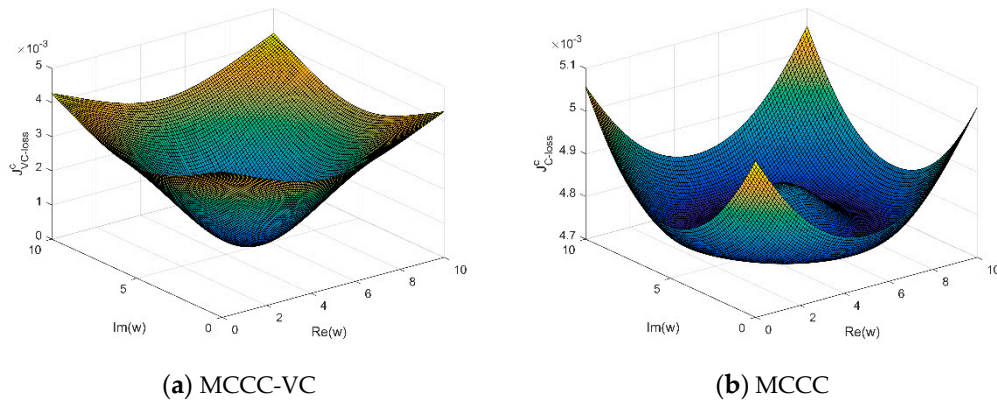


Figure 1. Surfaces of maximum complex correntropy criterion with variable center (MCCC-VC) and MCCC.

3.2. Gradient Descent Algorithm Based On MCCC-VC

Since the stochastic gradient descent approach requires less computational complexity, we adopt it to search the minimum of the cost function. Utilizing Wirtinger Calculus [27,28], we obtain the updation of the weight as follows:

$$\begin{aligned}
 \mathbf{w}(k+1) &= \mathbf{w}(k) - \mu \left\{ \frac{\partial \left[1 - \exp \left[-\frac{(e(k)-c(k))(e(k)-c(k))^*}{2\sigma^2} \right] \right]}{\partial \mathbf{w}^*(k)} \right\} \\
 &= \mathbf{w}(k) + \frac{\mu}{2\sigma^2} \exp \left[-\frac{|e(k)-c(k)|^2}{2\sigma^2} \right] (e(k) - c(k))^* \mathbf{x}(k) \\
 &= \mathbf{w}(k) + \eta_w \exp \left[-\frac{|e(k)-c(k)|^2}{2\sigma^2} \right] (e(k) - c(k))^* \mathbf{x}(k)
 \end{aligned} \tag{10}$$

where $\eta_w = \frac{\mu}{2\sigma^2}$ is the learning rate for the weight.

3.3. Optimization of the Parameters in MCCC-VC

3.3.1. Optimization Problem in MCCC-VC

The two parameters center location c and the width of kernel σ act a pivotal part in the performance of MCCC-VC. Thus, it is extremely important to optimize them to further improve the robustness and convergence performance in the non-zero mean noise.

The optimal model according to MCCC-VC is as follows:

$$M^* = \operatorname{argmax}_{M \in \mathcal{M}, \sigma \in \Omega, c \in \mathbb{C}} V_{\sigma, c}^C(T, Y) = \operatorname{argmax}_{M \in \mathcal{M}, \sigma \in \Omega, c \in \mathbb{C}} E[G_{\sigma}^C(e - c)] \quad (11)$$

In addition, the complex correntropy with variable center can be divided into three parts:

$$\begin{aligned} V_{\sigma, c}^C(T, Y) &= \iint G_{\sigma}^C(\varepsilon_R - c_R, \varepsilon_I - c_I) p_e(\varepsilon_R, \varepsilon_I) d\varepsilon_R d\varepsilon_I \\ &= \frac{1}{2} \iint [G_{\sigma}^C(\varepsilon_R - c_R, \varepsilon_I - c_I)]^2 d\varepsilon_R d\varepsilon_I + \frac{1}{2} \iint [p_e(\varepsilon_R, \varepsilon_I)]^2 d\varepsilon_R d\varepsilon_I \\ &\quad - \frac{1}{2} \iint [G_{\sigma}^C(\varepsilon_R - c_R, \varepsilon_I - c_I) - p_e(\varepsilon_R, \varepsilon_I)]^2 d\varepsilon_R d\varepsilon_I \end{aligned} \quad (12)$$

Owing to the first term is independent from the optimal model, we can derive

$$M^* = \operatorname{argmax}_{M \in \mathcal{M}, \sigma \in \Omega, c \in \mathbb{C}} V_{\sigma, c}^C(T, Y) = \operatorname{argmax}_{M \in \mathcal{M}, \sigma \in \Omega, c \in \mathbb{C}} U_{\sigma, c}^C(T, Y) \quad (13)$$

where

$$\begin{aligned} U_{\sigma, c}^C(T, Y) &= \iint [p_e(\varepsilon_R, \varepsilon_I)]^2 d\varepsilon_R d\varepsilon_I \\ &\quad - \iint [G_{\sigma}^C(\varepsilon_R - c_R, \varepsilon_I - c_I) - p_e(\varepsilon_R, \varepsilon_I)]^2 d\varepsilon_R d\varepsilon_I \end{aligned} \quad (14)$$

and

$$p_e(\varepsilon_R, \varepsilon_I) = 2E[G_{\sigma}^C(e_R - c_R, e_I - c_I)] \quad (15)$$

The parameters can be optimized by

$$(M^*, \sigma^*, c^*) = \operatorname{argmax}_{M \in \mathcal{M}, \sigma \in \Omega, c \in \mathbb{C}} U_{\sigma, c}^C(T, Y) \quad (16)$$

where Ω and \mathbb{C} represent the allowed sets of parameters σ and c .

Remark 1. It can be seen that as long as the function $U_{\sigma, c}^C(T, Y)$ is maximized, M , σ , and c can be optimized simultaneously. However, it is computationally demanding to compute and compare all the values of $U_{\sigma, c}^C(T, Y)$ under all the possible parameters in the allowed sets. Moreover, it may be difficult to obtain the allowed sets of parameters.

3.3.2. Stochastic Gradient Descent Approach

To further simplify the optimization problem, we propose a stochastic gradient descent based online approach.

(1) When the model M is fixed, $\iint p_e(\varepsilon) d\varepsilon_R d\varepsilon_I$ is independent of the kernel width σ and the center position c . In this case, σ and c can be optimized according to the following formula:

$$\begin{aligned} (\sigma^*, c^*) &= \operatorname{argmin}_{\sigma \in \Omega, c \in \mathbb{C}} \iint [G_{\sigma}^C(\varepsilon_R - c_R, \varepsilon_I - c_I) - p_e(\varepsilon_R, \varepsilon_I)]^2 d\varepsilon_R d\varepsilon_I \\ &= \operatorname{argmin}_{\sigma \in \Omega, c \in \mathbb{C}} \left\{ \iint [G_{\sigma}^C(\varepsilon_R - c_R, \varepsilon_I - c_I)]^2 d\varepsilon_R d\varepsilon_I - 2E[G_{\sigma}^C(e - c)] \right\} \\ &= \operatorname{argmin}_{\sigma \in \Omega, c \in \mathbb{C}} \left\{ -2E[G_{\sigma}^C(e - c)] + \frac{1}{4\pi\sigma^2} \right\} \end{aligned} \quad (17)$$

Provide N error samples $\{e(k)\}_{k=1}^N$, we can get $E[G_\sigma^C(e - c)] \approx \frac{1}{N} \sum_{k=1}^N G_\sigma^C(e(k) - c(k))$. Therefore, we have the following formula:

$$(\sigma^*, c^*) = \underset{\sigma \in \Omega, c \in \mathbb{C}}{\operatorname{argmin}} \left\{ - \left[\frac{2}{N} \sum_{k=1}^N G_\sigma^C(e(k) - c(k)) \right] + \frac{1}{4\pi\sigma^2} \right\} \quad (18)$$

Furthermore, in order to simplify the optimization problem, we can set $c(k)$ as the median or mean of the error samples. Thus, we only need to optimize σ . We take $1/\sigma^2$ as a new variable $\tilde{\sigma}$, and update $\tilde{\sigma}$ and σ^2 using the stochastic gradient descent approach as follows:

$$\begin{aligned} \tilde{\sigma}(k+1) &= \tilde{\sigma}(k) - \eta_\sigma \left[\frac{\partial - \left[\frac{2}{N} \sum_{l=k-T+1}^k G_\sigma^C(e(l) - c(k)) \right] + \frac{1}{4\pi\sigma^2}}{\partial \tilde{\sigma}} \right] \Big|_{\tilde{\sigma}=\tilde{\sigma}(k), c=c(k)} \\ &= \tilde{\sigma}(k) - \eta_\sigma \left\{ - \left[\frac{1}{\pi N} \sum_{l=k-T+1}^k \exp \left(-\frac{|e(l) - c(k)|^2}{2} \tilde{\sigma}(k) \right) \left(1 - \frac{|e(l) - c(k)|^2}{2} \tilde{\sigma}(k) \right) \right] + \frac{1}{4\pi} \right\} \end{aligned} \quad (19)$$

and

$$\sigma^2(k+1) = \frac{1}{\tilde{\sigma}(k+1)} \quad (20)$$

where $c(k)$ is estimated online as $c(k) = \sum_{l=k-T+1}^k e(l)$, and T is the smoothing length, η_σ is the learning rate for $\tilde{\sigma}$.

(2) When the kernel width $\sigma(k)$ and the center position $c(k)$ is fixed, the model M is optimized by MCCC-VC using (10).

Remark 2. For the proposed MCCC-VC algorithm, the weight and the parameters are updated alternately at each time instant k using (10), (19) and (20), respectively.

3.4. Performance Analysis

3.4.1. Convergence Analysis

The MCCC-VC algorithm is written as a form of nonlinear error:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \eta_w f(e(k)) \mathbf{x}(k) \quad (21)$$

with $f(e(k)) = \exp \left[-\frac{|e(k) - c(k)|^2}{2\sigma^2} \right] (e(k) - c(k))^*$ being the scalar function of the error $e(k)$.

Taking into consideration that

$$d(k) = \mathbf{w}_0^H \mathbf{x}(k) + v(k) \quad (22)$$

the error is written as

$$e(k) = \tilde{\mathbf{w}}^H(k) \mathbf{x}(k) + v(k) = e_a(k) + v(k) \quad (23)$$

where $\tilde{\mathbf{w}}(k) = \mathbf{w}_0 - \mathbf{w}(k)$ is the weight error vector at time instant k , \mathbf{w}_0 is the system parameter, $e_a(k) = \tilde{\mathbf{w}}^H(k) \mathbf{x}(k)$ is the prior error, and $v(k)$ is the additive noise at time instant k .

Therefore, we get the following formula

$$\tilde{\mathbf{w}}(k+1) = \tilde{\mathbf{w}}(k) - \eta_w f(e(k)) \mathbf{x}(k) \quad (24)$$

By taking the square of the 2-norm of both sides, we can further get the following formula:

$$\begin{aligned} & E\{\|\tilde{\mathbf{w}}(k+1)\|^2\} \\ &= E\{\|\tilde{\mathbf{w}}(k)\|^2\} - 2\eta_w E\{\text{Re}[e_a(k)f(e(k))]\} \\ & \quad + \eta_w^2 E\{\|\mathbf{x}(k)\|^2 |f(e(k))|^2\} \end{aligned} \quad (25)$$

To guarantee the convergence of the MCCC-VC, the weight error power should be gradually decreased. Thus, we obtain the bound for the learning rate as follows:

$$0 < \eta_w \leq \frac{2E\{\text{Re}[e_a(k)f(e(k))]\}}{E\{\|\mathbf{x}(k)\|^2 |f(e(k))|^2\}} \quad (26)$$

3.4.2. Steady-State Mean Square

If MCCC-VC arrives at steady-state, we have

$$\lim_{k \rightarrow \infty} E\{\|\tilde{\mathbf{w}}(k+1)\|^2\} = \lim_{k \rightarrow \infty} E\{\|\tilde{\mathbf{w}}(k)\|^2\} \quad (27)$$

Then, when $k \rightarrow \infty$, we can get

$$2E\{\text{Re}[(e_a(k) - c)f(e(k))]\} = \eta_w E\{\|\mathbf{x}(k)\|^2 |f(e(k))|^2\} \quad (28)$$

According to the definition of the steady-state excess mean square error (EMSE), we have

$$S = \lim_{k \rightarrow \infty} E[|e_a(k)|^2] = E[|e_a|^2] \quad (29)$$

To obtain the theoretical steady-state EMSE, we present the following two assumptions [21,22,29]:

- (1) $v(k)$ is zero-mean distributed and independent of $x(k)$, and $x(k)$ is circular.
- (2) $e_a(k)$ is zero-mean and independent of $v(k)$.

Owing to the distributions of $x(k)$, $v(k)$, $e_a(k)$, and $e(k)$ are not related to the time index k at the steady-state, the time index is ignored in the following derivation.

The left side of (28) can be written as

$$\begin{aligned} L &= E\left\{e_a \left[\exp\left[-\frac{|e-c|^2}{2\sigma^2}\right] (e-c)^* + e_a^* \left[\exp\left[-\frac{|e-c|^2}{2\sigma^2}\right] (e-c)\right]\right\} \\ &= E\left\{\exp\left[-\frac{|e-c|^2}{2\sigma^2}\right] (e_a(e-c)^* + e_a^*(e-c))\right\} \\ &= E\{g_1(e)(2|e_a|^2 + e_a v^* + e_a^* v)\} \end{aligned} \quad (30)$$

where

$$g_1(e) = \exp\left[-\frac{|e-c|^2}{2\sigma^2}\right] \quad (31)$$

We use Taylor expansion to approximate $g_1(e)$ as

$$g_1(e) \approx g_1(v) + 2\text{Re}\left\{\frac{\partial g_1}{\partial e}\bigg|_{e=v} \cdot e_a\right\} + \text{Re}\left\{\frac{\partial^2 g_1}{\partial e^* \partial e}\bigg|_{e=v} \cdot (e_a^*)^2 + \frac{\partial^2 g_1}{\partial e^* \partial e}\bigg|_{e=v} \cdot |e_a|^2\right\} \quad (32)$$

where

$$\frac{\partial g_1}{\partial e} = \exp\left[-\frac{|e-c|^2}{2\sigma^2}\right] \times \left[-\frac{|e-c|^2}{2\sigma^2} (e-c)^{-1}\right] \quad (33)$$

$$\frac{\partial g_1}{\partial e^*} = \exp\left[-\frac{|e-c|^2}{2\sigma^2}\right] \times \left[-\frac{|e-c|^2}{2\sigma^2} ((e-c)^*)^{-1}\right] \quad (34)$$

$$\frac{\partial^2 g_1}{\partial e^* \partial e} = \exp\left[-\frac{|e-c|^2}{2\sigma^2}\right] |e-c|^{-2} \times \left[\frac{|e-c|^4}{(2\sigma^2)^2} - \frac{3|e-c|^2}{2\sigma^2} + \frac{|e-c|^2}{\sigma^2}\right] \quad (35)$$

$$\frac{\partial^2 g_1}{\partial e^* \partial e^*} = \exp\left[-\frac{|e-c|^2}{2\sigma^2}\right] ((e-c)^*)^{-2} \times \left[\frac{|e-c|^4}{(2\sigma^2)^2}\right] \quad (36)$$

Owing to x is circular, we can get the values of the following two items:

$$E[(e_a^*)^2] = 0 \quad (37)$$

$$E[e_a^2] = \tilde{\mathbf{w}}^H \mathbf{x} \mathbf{x}^T \tilde{\mathbf{w}}^* = 0 \quad (38)$$

Based on the above derivation, if the higher-order terms are small enough, we can rewrite the left side of (28) as follows:

$$L \approx 2S \exp\left[-\frac{|v|^2}{2\sigma^2}\right] \times \left\{1 - \frac{|v|^2}{2\sigma^2}\right\} \quad (39)$$

The right side of (28) can be written as

$$\begin{aligned} R &= \eta_w \text{Tr}(\mathbf{R}_{\mathbf{x}\mathbf{x}^H}) E\left\{|f(e(k))|^2\right\} \\ &= \eta_w \text{Tr}(\mathbf{R}_{\mathbf{x}\mathbf{x}^H}) E\left\{\exp\left[-\frac{|e-c|^2}{\sigma^2}\right] |e-c|^2\right\} \\ &= \eta_w \text{Tr}(\mathbf{R}_{\mathbf{x}\mathbf{x}^H}) E\{g_2(e)\} \end{aligned} \quad (40)$$

where

$$g_2(e) = \exp\left[-\frac{|e-c|^2}{\sigma^2}\right] |e-c|^2 \quad (41)$$

In a similar way, we use a Taylor expansion to approximate $g_2(e)$ as

$$g_2(e) \approx g_2(v) + \text{Re}\left\{\frac{\partial^2 g_2}{\partial e^* \partial e}\bigg|_{e=v} \cdot |e_a|^2 + \frac{\partial^2 g_2}{\partial e^* \partial e^*}\bigg|_{e=v} \cdot (e_a^*)^2\right\} + 2\text{Re}\left\{\frac{\partial g_2}{\partial e}\bigg|_{e=v} \cdot e_a\right\} \quad (42)$$

where

$$\frac{\partial g_2}{\partial e} = \exp\left[-\frac{|e-c|^2}{\sigma^2}\right] |e-c|^2 \times \left[-\frac{|e-c|^2}{\sigma^2} ((e-c)^*)^{-1} + (e^*)^{-1}\right] \quad (43)$$

$$\frac{\partial g_2}{\partial e^*} = \exp\left[-\frac{|e-c|^2}{\sigma^2}\right] |e-c|^2 \times \left[-\frac{|e-c|^2}{\sigma^2} ((e-c)^*)^{-1} + (e^*)^{-1}\right] \quad (44)$$

$$\frac{\partial^2 g_2}{\partial e^* \partial e} = \exp\left[-\frac{|e-c|^2}{\sigma^2}\right] \times \left[\frac{|e-c|^4}{\sigma^4} - \frac{3|e-c|^2}{\sigma^2} - 1\right] \quad (45)$$

$$\frac{\partial^2 g_2}{\partial e^* \partial e^*} = \exp\left[-\frac{|e-c|^2}{2\sigma^2}\right] |e-c|^2 ((e-c)^*)^{-2} \left(\frac{|e-c|^4}{\sigma^4} - \frac{|e-c|^2}{2\sigma^2}\right) \quad (46)$$

If the higher-order terms are small enough, we can rewrite the right side of (28) as follows:

$$R \approx \eta_w \text{Tr}(\mathbf{R}_{\mathbf{x}\mathbf{x}^H}) E\left\{\exp\left[-\frac{|v-c|^2}{2\sigma^2}\right] |v-c|^2\right\} + \eta_w \text{Tr}(\mathbf{R}_{\mathbf{x}\mathbf{x}^H}) S \times R_1 \quad (47)$$

where

$$R_1 = E\left\{\exp\left[-\frac{|v-c|^2}{2\sigma^2}\right] \left(\frac{|v-c|^4}{\sigma^4} - \frac{3|v-c|^2}{\sigma^2} - 1\right)\right\} \quad (48)$$

Finally, we get the theoretical steady-state EMSE as follows:

$$S = \frac{\eta_w \text{Tr}(\mathbf{R}_{\mathbf{x}\mathbf{x}^H}) E\left\{\exp\left[-\frac{|v-c|^2}{2\sigma^2}\right] |v-c|^2\right\}}{E\left\{2 \exp\left[-\frac{|v-c|^2}{2\sigma^2}\right] \times \left[1 - \frac{|v-c|^2}{2\sigma^2}\right]\right\} - \eta_w \text{Tr}(\mathbf{R}_{\mathbf{x}\mathbf{x}^H}) R_1} \quad (49)$$

Furthermore, when η_w is small enough, (49) is further simplified as

$$S = \frac{\eta_w \text{Tr}(\mathbf{R}_{\mathbf{x}\mathbf{x}^H}) E\left\{\exp\left[-\frac{|v-c|^2}{2\sigma^2}\right] |v-c|^2\right\}}{E\left\{2 \exp\left[-\frac{|v-c|^2}{2\sigma^2}\right] \times \left[1 - \frac{|v-c|^2}{2\sigma^2}\right]\right\}} \quad (50)$$

Moreover, we derive the theoretical value of σ^2 by setting $\frac{\partial\{-2E[G_\sigma^C(e-c)] + \frac{1}{4\pi\sigma^2}\}}{\partial\sigma^2} = 0$. In this way, we have $\frac{1}{\pi\sigma^4} \exp\left(-\frac{|e-c|^2}{2\sigma^2}\right) - \frac{1}{\pi\sigma^6} \exp\left(-\frac{|e-c|^2}{2\sigma^2}\right) \frac{|e-c|^2}{2} - \frac{1}{4\pi\sigma^4} = 0$. Due to $e \approx v$ at the steady state, we can further obtain the theoretical value of σ^2 based on the following approach:

$$\sigma^2 = \frac{E\left\{\frac{|v-c|^2}{2} \exp\left[-\frac{|v-c|^2}{2\sigma^2}\right]\right\}}{E\left\{-\frac{1}{4} + \exp\left[-\frac{|v-c|^2}{2\sigma^2}\right]\right\}} \quad (51)$$

Since the right side of (51) depends on σ^2 , it is a fixed-point solution for the theoretical σ^2 .

Remark 3. The theoretical steady-state EMSE in (50) is accurate only when e_a is small enough, since the higher-order term can be negligible in this case. However, if the noise power or step size is too large, or the center position of the kernel function deviates from the mean of the noise, there will be a large deviation between the theoretical and simulated values of steady-state EMSE.

4. Simulation

In this section, we present some simulations to show the validity of theoretical results and the superiority of MCCC-VC. We obtain all the simulation results by averaging over 300 Monte Carlo trials.

4.1. Steady-State Performance

In this part, the filter weight $\mathbf{w}_0 = [w_1 \ w_2 \ \cdots \ w_{10}]^T$ is randomly generated, where $w_k = w_{Rk} + jw_{Ik}$, and $w_{Rk}, w_{Ik} \in N(0, 0.1)$, w_{Rk} and w_{Ik} represent the real and imaginary components of w_k , and $N(\mu, \hat{\sigma}^2)$ denotes the Gaussian distributed variable whose mean and variance are μ and $\hat{\sigma}^2$,

respectively. We randomly generate input signal $x = x_R + jx_I$. In order to show the robustness of MCCC-VC, additive complex noise $v = v_R + jv_I$ is added in the simulation, whose real and imaginary parts are denoted by v_R and v_I , respectively. All algorithms initialize \mathbf{w} with a zero vector.

Firstly, we illustrate the correctness of theoretical steady-state EMSEs. For each simulation, 30,000 iterations are carried out to make sure MCCC-VC reaches the steady-state, and the last 1000 iterations are used to obtain the simulated steady-state EMSEs. The theoretical kernel width and steady-state EMSEs are calculated according to (51) and (50), respectively. Figures 2 and 3 show the simulated and theoretical steady-state EMSEs of MCCC-VC under various noise variances and learning rates, where v is Gaussian distributed with mean $3 + 3j$. It can be seen from both figures that theoretical results are closely matching with simulated results.

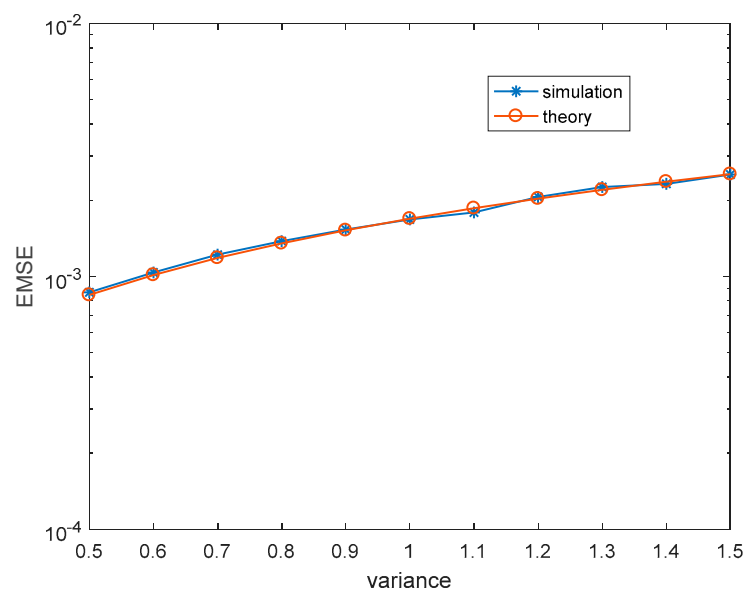


Figure 2. Steady-state excess mean square errors (EMSEs) under various σ_v^2 . (Gaussian distributed noise, $\eta_w = 3.8 \times 10^{-4}$, $\eta_\sigma = 4 \times 10^{-3}$).

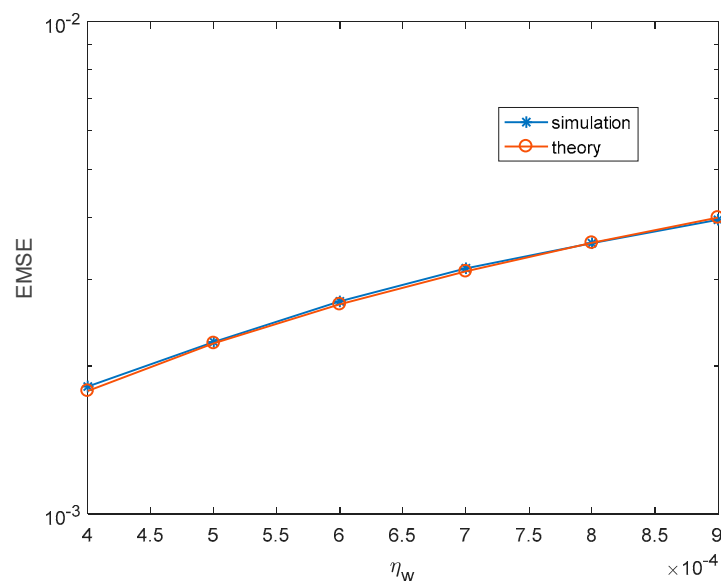


Figure 3. Steady-state EMSEs under various η_w . (Gaussian distributed noise, $\sigma_v^2 = 1$, $\eta_\sigma = 4 \times 10^{-3}$).

Then, we change the noise to binary noise, and the mean is also $3 + 3j$. In addition, the simulated and theoretical steady-state EMSEs are obtained the same as before. Figures 4 and 5 show the simulated

and theoretical steady-state results of MCCC-VC under various noise variances and learning rates. Obviously, there is also a good matching between theoretical results and simulated results.

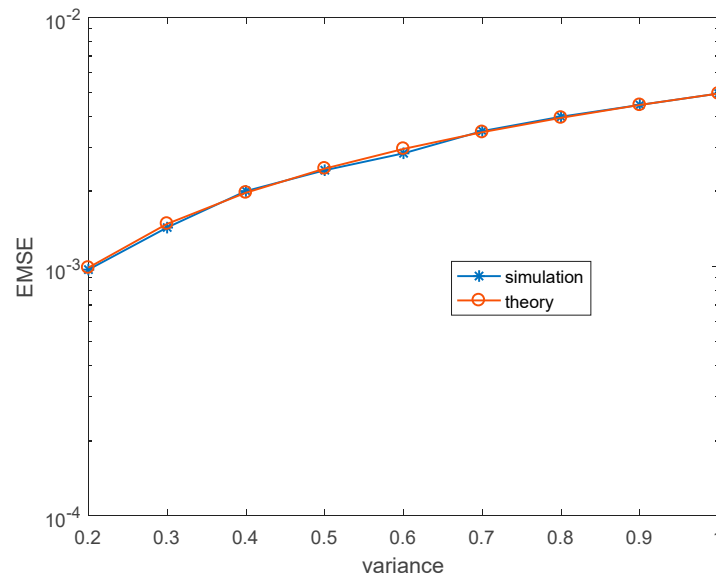


Figure 4. Steady-state EMSEs under various σ_v^2 . (Binary distributed noise, $\eta_w = 3.8 \times 10^{-4}$, $\eta_\sigma = 4 \times 10^{-3}$).

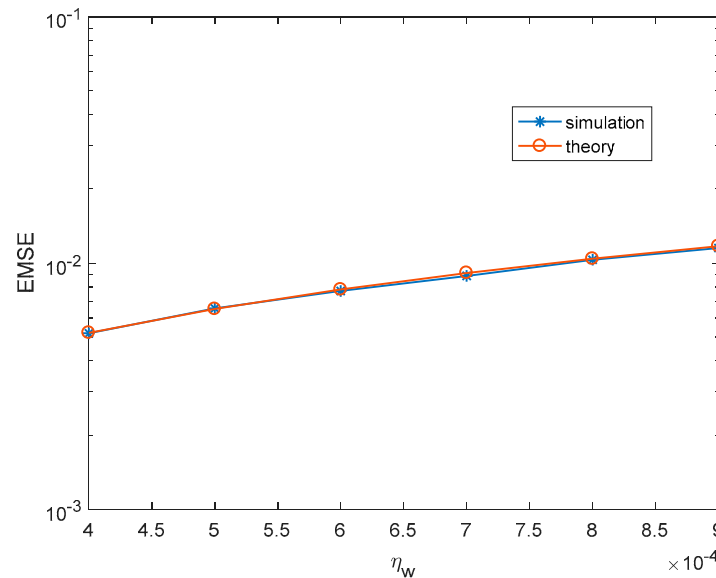


Figure 5. Steady-state EMSEs under various η_w . (Binary distributed noise, $\sigma_v^2 = 1$, $\eta_\sigma = 4 \times 10^{-3}$).

4.2. Performance Comparison

In this part, we compare the performance of the proposed MCCC-VC algorithm with MCCC and minimum complex kernel risk sensitive loss (MCKRSL) [22]. For the fair comparison, all three algorithms use the gradient descent method to search for the optimal solution. We measure the performance of all the algorithms by weight error power.

In this simulation, the noise $v(k)$ is composed of two independent noises [16], i.e., $v(k) = (1 - a(k))A(k) + a(k)B(k)$, where $P(a(k) = 0) = 1 - c$, and $P(a(k) = 1) = c$ ($0 \leq c \leq 1$). $A(k)$ is the ordinary noise with small variance $\sigma_v^2 = 1$ whose real and imaginary parts are denoted by $A_R(k)$ and $A_I(k)$, and $B(k)$ is the outliers with large variance whose real and imaginary parts are denoted by $B_R(k)$ and $B_I(k)$.

In this simulation, we set $c = 0.05$ and $B_R, B_I \in N(0, 100)$. In addition, we consider the following four cases for $A(k)$:

- (1) $A_R(k), A_I(k) \in N(3, \sigma_v^2/2)$;
- (2) $P(A_R(k) = 3 + \sigma_v/\sqrt{2}) = P(A_R(k) = 3 - \sigma_v/\sqrt{2}) = P(A_I(k) = 3 + \sigma_v/\sqrt{2}) = P(A_I(k) = 3 - \sigma_v/\sqrt{2}) = 0.5$;
- (3) $A_R(k), A_I(k) \in U(3 - \sigma_v/\sqrt{2}, 3 + \sigma_v/\sqrt{2})$, with $U(\alpha, \beta)$ denoting the uniform distribution over $[\alpha, \beta]$;
- (4) $A_R(k) = 3 + \sigma_v \sin \theta_{1k}/\sqrt{2}$, $A_I(k) = 3 + \sigma_v \sin \theta_{2k}/\sqrt{2}$, where $\theta_{1k}, \theta_{2k} \in U[0, 2\pi]$.

Figures 6–9 show the convergence behavior of various algorithms on the basis of weight error power $\|\mathbf{w}(k) - \mathbf{w}_0\|^2$ under different noises, where the parameter settings of different algorithms are summarized in Table 1. It can be seen clearly that the convergence performance of MCCC-VC is better than other two algorithms in all cases.

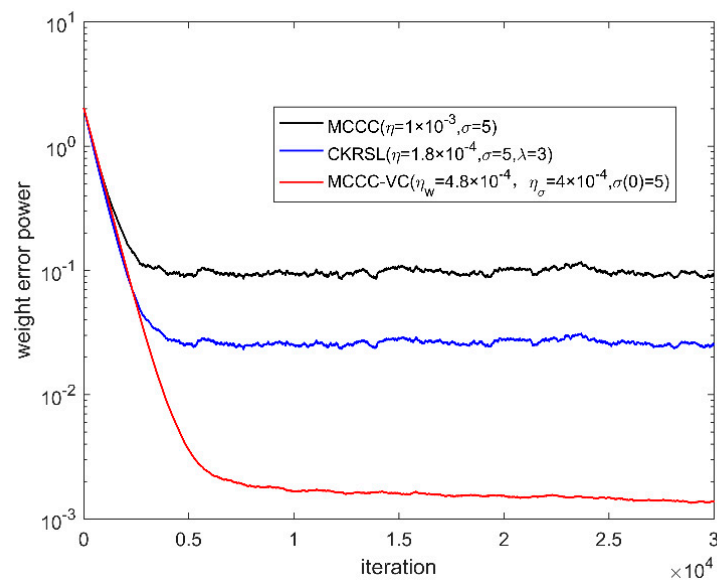


Figure 6. Convergence behavior of various algorithms (case 1).

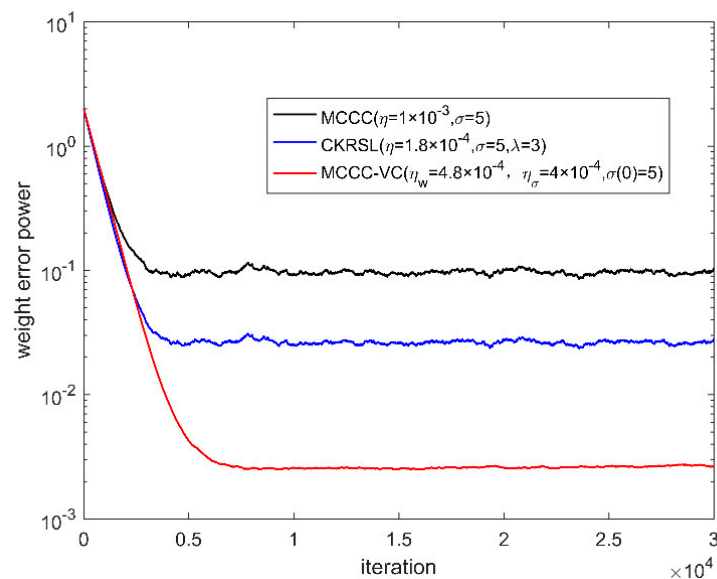


Figure 7. Convergence behavior of various algorithms (case 2).

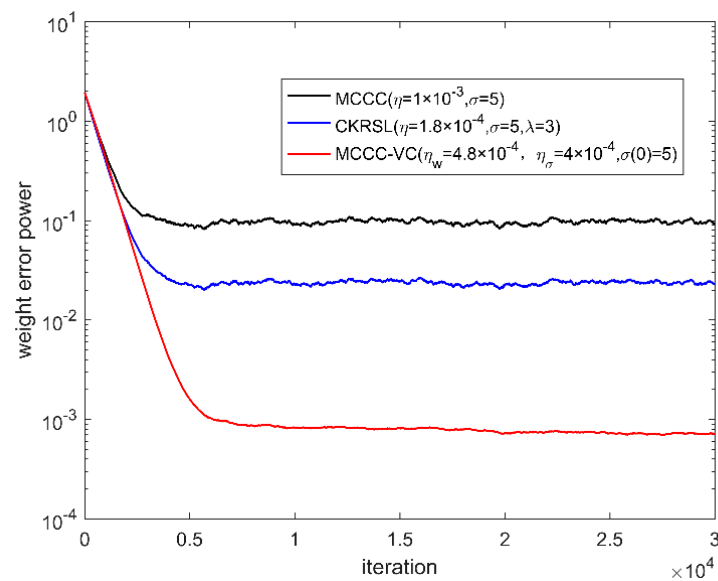


Figure 8. Convergence behavior of various algorithms (case 3).

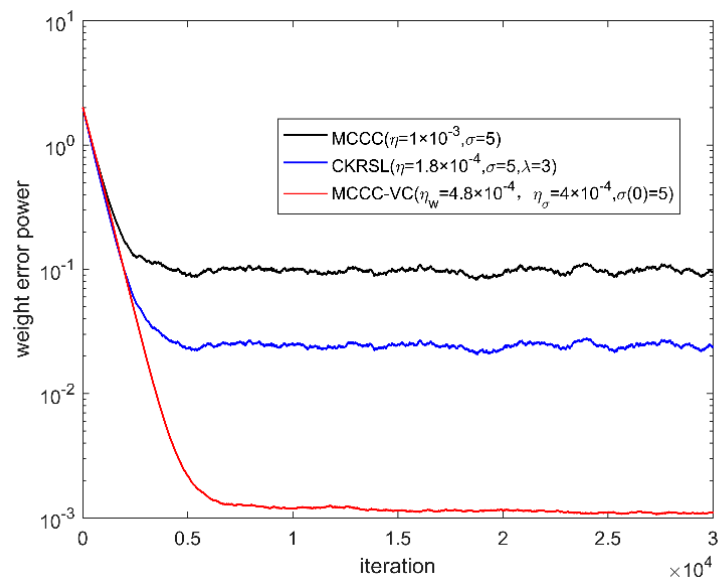


Figure 9. Convergence behavior of various algorithms (case 4).

Table 1. Parameter setting of different algorithms.

Algorithm	MCCC	MCKRSL	MCCC-VC
Parameters	$\eta = 1 \times 10^{-3},$ $\sigma = 5.$	$\eta = 1.8 \times 10^{-4},$ $\sigma = 5,$ $\lambda = 3.$	$\eta_w = 4.8 \times 10^{-4},$ $\eta_\sigma = 4 \times 10^{-4},$ $\sigma(0) = 5.$

Notes: η and σ denote the learning rate and kernel width for MCCC and MCKRSL, and λ denotes the risk-sensitive parameter for MCKRSL. Moreover, η_w, η_σ denote the learning rates for the weight and kernel width of MCCC-VC, and $\sigma(0)$ denotes the initial kernel width of MCCC-VC.

5. Conclusions

The complex correntropy usually employs a Gaussian kernel whose center is zero, which is not the best choice for many situations. To overcome this defect, this paper proposes the maximum complex correntropy criterion with variable center (MCCC-VC). The complex correntropy is extended to the case where the center can be anywhere. Furthermore, this paper also proposes an effective method to

optimize the center position and the kernel width. More significantly, we analyze the convergence and steady-state performance of MCCC-VC theoretically. Simulation results obtained in Section 4 support the reliability of theoretical analysis and show the excellent performance of MCCC-VC.

Author Contributions: Conceptualization, F.D., G.Q., and S.W.; methodology, F.D., G.Q., and S.W.; software, F.D., G.Q., and S.W.; validation, G.Q.; formal analysis, G.Q.; investigation, F.D., G.Q., and S.W.; resources, G.Q.; data curation, F.D. and G.Q.; writing—original draft preparation, F.D.; writing—review and editing, G.Q. and S.W.; visualization, F.D. and G.Q.; supervision, F.D., G.Q., and S.W.; project administration, G.Q.; funding acquisition, S.W. and G.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China under grants 61671389 and 61701419, and Fundamental Research Funds for the Central Universities under grant XDJK2019B011.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Haykin, S. *Adaptive Filter Theory*, 3rd ed.; Prentice Hall: New York, NY, USA, 1996.
- Sayed, A.H. Fundamentals of adaptive filtering. *IEEE Control Syst.* **2003**, *25*, 77–79.
- Chen, B.; Zhu, Y.; Hu, J.; Principe, J.C. *System Parameter Identification: Information Criteria and Algorithms*; Newnes: Oxford, UK, 2013.
- Widrow, B.; McCool, J.M.; Larimore, M.G.; Johnson, C.R. Stationary and nonstationary learning characteristics of the LMS adaptive filter. *Proc. IEEE*. **1976**, *64*, 1151–1162. [[CrossRef](#)]
- Kwong, R.H.; Johnston, E.W. A variable step size LMS algorithm. *IEEE Trans. Signal Process.* **1992**, *40*, 1633–1642. [[CrossRef](#)]
- Benesty, J.; Duhamel, P. A fast exact least mean square adaptive algorithm. *IEEE Trans. Signal Process.* **1992**, *40*, 2904–2920. [[CrossRef](#)]
- Diniz, P.S.R. *Adaptive Filtering: Algorithms and Practical Implementation*, 4th ed.; Springer: New York, NY, USA, 2013.
- Pei, S.C.; Tseng, C.C. Least mean p-power error criterion for adaptive FIR filter. *IEEE J. Sel. Areas Commun.* **1994**, *12*, 1540–1547.
- Al-Naffouri, T.Y.; Sayed, A.H. Adaptive filters with error nonlinearities: Mean-square analysis and optimum design. *EURASIP J. Appl. Signal Process.* **2001**, *1*, 192–205. [[CrossRef](#)]
- Erdogmus, D.; Principe, J.C. Generalized information potential criterion for adaptive system training. *IEEE Trans. Neural Netw.* **2002**, *13*, 1035–1044. [[CrossRef](#)]
- Principe, J.C. *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*; Springer: New York, NY, USA, 2010.
- Sayin, M.O.; Vanli, N.D.; Kozat, S.S. A novel family of adaptive filtering algorithms based on the logarithmic cost. *IEEE Trans. Signal Process.* **2014**, *62*, 4411–4424. [[CrossRef](#)]
- Liu, W.; Pokharel, P.P.; Principe, J.C. Correntropy: Properties and applications in non-Gaussian signal processing. *IEEE Trans. Signal Process.* **2007**, *55*, 5286–5298. [[CrossRef](#)]
- Chen, B.; Xing, L.; Zhao, H.; Zheng, N.; Principe, J.C. Generalized correntropy for robust adaptive filtering. *IEEE Trans. Signal Process.* **2016**, *64*, 3376–3387. [[CrossRef](#)]
- Ma, W.; Qu, H.; Gui, G.; Xu, L.; Zhao, J.; Chen, B. Maximum correntropy criterion based sparse adaptive filtering algorithms for robust channel estimation under non-Gaussian environments. *J. Franklin Inst.* **2015**, *352*, 2708–2727. [[CrossRef](#)]
- Chen, B.; Xing, L.; Xu, B.; Zhao, H.; Zheng, N.; Principe, J.C. Kernel Risk-Sensitive Loss: Definition, Properties and Application to Robust Adaptive Filtering. *IEEE Trans. Signal Process.* **2017**, *65*, 2888–2901. [[CrossRef](#)]
- Mandic, D.; Goh, V. *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models* (ser. *Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control*); Wiley: New York, NY, USA, 2009.
- Shi, L.; Zhao, H.; Zakharov, Y. Performance analysis of shrinkage linear complex-valued LMS algorithm. *IEEE Signal Process. Lett.* **2019**, *26*, 1202–1206. [[CrossRef](#)]
- Guimaraes, J.P.F.; Fontes, A.I.R.; Rego, J.B.A.; Martins, A.M.; Principe, J.C. Complex correntropy: Probabilistic interpretation and application to complex-valued data. *IEEE Signal Process. Lett.* **2017**, *24*, 42–45. [[CrossRef](#)]

20. Guimarães, J.P.F.; Fontes, A.I.R.; Rego, J.B.A.; Martins, A.M.; Principe, J.C. Complex Correntropy Function: properties, and application to a channel equalization problem. *Expert Syst. Appl.* **2018**, *107*, 173–181. [[CrossRef](#)]
21. Qian, G.; Wang, S. Generalized Complex Correntropy: Application to Adaptive Filtering of Complex Data. *IEEE Access.* **2018**, *6*, 19113–19120. [[CrossRef](#)]
22. Qian, G.; Wang, S. Complex Kernel Risk-Sensitive Loss: Application to Robust Adaptive Filtering in Complex Domain. *IEEE Access.* **2018**, *6*, 60329–60338. [[CrossRef](#)]
23. Guimarães, J.P.F.; Fontes, A.I.R.; da Silva, F.B. Complex Correntropy Induced Metric Applied to Compressive Sensing with Complex-Valued Data. In Proceedings of the 2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), Las Vegas, NV, USA, 8–10 April 2018; pp. 21–24.
24. Qian, G.; Luo, D.; Wang, S. A Robust Adaptive Filter for a Complex Hammerstein System. *Entropy* **2019**, *21*, 162. [[CrossRef](#)]
25. Chen, B.; Wang, X.; Li, Y.; Principe, J.C. Maximum correntropy criterion with variable center. *IEEE Signal Process.* **2019**, *26*, 1212–1216. [[CrossRef](#)]
26. Zhu, L.; Song, C.; Pan, L.; Li, J. Adaptive filtering under the maximum correntropy criterion with variable center. *IEEE Access.* **2019**, *7*, 105902–105908. [[CrossRef](#)]
27. Wirtinger, W. Zur formalen theorie der funktionen von mehr complexen veränderlichen. *Math. Ann.* **1927**, *97*, 357–375. [[CrossRef](#)]
28. Bouboulis, P.; Theodoridis, S. Extension of Wirtinger’s calculus to reproducing Kernel Hilbert spaces and the complex kernel LMS. *IEEE Trans. Signal Process.* **2011**, *59*, 964–978. [[CrossRef](#)]
29. Picinbono, B. On circularity. *IEEE Trans. Signal Process.* **1994**, *42*, 3473–3482. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).